

FRONT PAGE

PROJECT PERIODIC REPORT

Grant Agreement number: 223758

Project acronym: PARSE.INSIGHT

Project title: INSIGHT into issues of Permanent Access to the Records of Science in Europe

Funding Scheme: Support Action (CSA-SA)

Date of latest version of Annex I against which the assessment will be made: 5 Dec 2007

Periodic report: 1st ☒ 2nd ☐ 3rd ☐ 4th ☐

Period covered: from 1 March 2008 to 28 February 2009

Name, title and organisation of the scientific representative of the project's coordinator:

Dr David Giaretta
Science and Technology Facilities Council (STFC, UK)

Tel: +44 1235 446235

Fax: +44 1235 445848

E-mail: david.giaretta@stfc.ac.uk

Project website address: www.parse-insight.eu

1. Publishable summary



Objectives

The motivation of the PARSE.Insight project is to contribute to the long-term access to the digital resources created by scientific endeavour. It is widely recognised that there are risks that such resources might be lost to future use unless active steps are taken for their preservation. Not only hardware, media and formats change, but knowledge, required to interpret and reuse data, also changes over time. There are of course many initiatives under way dealing with this problem, but PARSE.Insight aims to look across communities to seek a common infrastructure.

The project aims to deliver:

- **Insight and understanding** into the capabilities and practices within the various research communities
- An **inventory** of current and planned research and development relating to e-infrastructures and permanent access
- A **roadmap** for a support e-infrastructure for maintaining long-term accessibility and usability of scientific and other digital information in Europe
- Identification of **gaps** in the existing and planned infrastructure
- Progress towards a standard for **evaluating** the sustainability and trustworthiness of digital repositories.

The PARSE.Insight project has strong links with the Alliance for Permanent Access to the Records of Science (<http://www.alliancepermanentaccess.eu>).

Work performed and results achieved

In the first year of project the main emphasis of the project has been surveying communities with an interest in digital preservation to build up insight, and developing a draft roadmap for the e-infrastructure.

General surveys and case studies have been performed. The general surveys were aimed at distinct groups of stakeholders: researchers, funders, publishers and data archivists. The surveys employed online questionnaires constructed to obtain information about the knowledge, attitudes, practices and desires of the stakeholders with respect to digital preservation. The communities were contacted using a wide range of mailing lists and other publicity.

| |
|---|
| See deliverables: D3.1 'Survey and forum platforms', D3.2 'Inventory of communities'. |
|---|

The general surveys have been one of the notable successes of the project in its first year. In total around 2000 responses were received, which provides an important base of evidence for the importance of digital preservation across a wide range of scientific disciplines and countries. The general surveys are being analysed and will be followed with interviews in the second year of the project to explore themes more deeply or to compensate for gaps in coverage.

One of the ideas underlying the surveys was *threats to preservation*—eventualities that could lead to the loss of digital resources or the inability to understand them. Between 50% and 70% of responses indicate that all the threats are recognized as either ‘Important’ or ‘Very Important’, with about half supporting the need for an international preservation infrastructure. Another clear message is that researchers would like to (re-)use data from both their own and other disciplines, and it is suggested that this is likely to produce more and better science. However more than 50% report that they have wished to access digital research data gathered by other researchers which turned out to be unavailable.

The case studies have a different motivation from the general survey, aiming to investigate more deeply and more narrowly the characteristics of certain communities. Four case studies are being conducted, in high energy physics, earth observation, psycholinguistics and book studies (the latter two being on a smaller scale). The approach in each is slightly different, depending on the characteristics of the community in question, but targeted surveys are being performed and it is expected that interviews will follow.

See deliverables: D3.3 ‘Case study reports’, D3.4 ‘Survey report’—both expected soon.

The draft roadmap is another major achievement of the first year of the project. Using the survey results as a foundation, the roadmap characterises what is meant by an e-infrastructure for science data and proposes components to build it. These components are financial, organisational/social, policy and technical. Technical components are the solutions to the threats to preservation. They have been illustrated with scenarios to show their relation to real needs.

See deliverable: D2.1 ‘Draft roadmap’.

During the course of the year, it was proposed to broaden the project scope away from preservation towards a more general science data infrastructure. One consequence of this is the elimination of the work on impact analysis, and its replacement by broadening and deepening the community insight and on a more qualitative approach to illustrating the consequences of the roadmap.

The gap analysis examines the gap between the current or foreseen situation in terms of developments in preservation, and the ideal that is envisaged in the roadmap. A systematic approach has been defined. First a gap analysis framework has been developed, eliciting and structuring the relevant dimensions and corresponding attributes for future e-infrastructures into a formal schema. Then a stepwise procedure was developed for conducting the gap analyses, providing methods and metrics to identify gaps within the European research infrastructure. This will be applied in the second year of the project.

See deliverable: D4.1 ‘Gap analysis framework specification’.

In the sustainability and evaluation work, the focus has been on the progress towards an international standard for audit and certification of digital repositories. A workshop was held in the

US at which excellent progress was made on the draft standard, which is now close to submission to the ISO process.

| |
|--|
| See deliverable: D6.1 'Workshop report'. |
|--|

There has been a large amount of publicity and dissemination of the project's work, with PARSE.Insight partners present at conferences giving talks and having papers published. Several workshops are being planned to engage important stakeholders in the remaining work of the project.

| |
|---|
| See deliverable: D7.1 'Dissemination plan'. |
|---|

Expected final results and their potential impact and use

By the end of the project an important base of data will have been assembled concerning the attitudes and practices of a wide range of scientific communities concerning digital preservation and science data infrastructure. This will provide an excellent body of evidence for policy makers, strategists and funders. The data will be both broad and deep (from the interviews and case studies).

The project will revise its roadmap, which will influence the agenda of development in the science data infrastructure for the coming years. The roadmap will be complemented by an understanding of the gaps with respect to the current situation. Additional stakeholders will be involved, from the Alliance for Permanent Access and through the series of workshops that are to be organised.

Find out more

The public website of the PARSE.Insight project is at <http://www.parse-insight.eu> (screen shot of the front page below). There the project's publications can be found, together with constantly updated news.

Permanent Access to the Records of Science in Europe



PARSE.Insight is a two-year project co-funded by the European Union under the Seventh Framework Programme. It is concerned with the preservation of digital information in science, from primary data through analysis to the final publications resulting from the research. The problem is how to safeguard this valuable digital material over time, to ensure that it is accessible, usable and understandable in future. The rapid pace of change in information technology threatens media, file formats and software with obsolescence, and changing concepts and terminology also mean that, even if data can be read, it might not be correctly interpreted by future generations.

Many initiatives are already under way in this area, and the aim of the PARSE.Insight project is to develop a roadmap and recommendations for developing the e-infrastructure in order to maintain the long-term accessibility and usability of scientific digital information in Europe. The project will conduct surveys and in-depth case studies of different scientific disciplines and stakeholders and will base its results on these findings, as well as knowledge of ongoing developments.

PARSE.Insight is closely linked to the Alliance for Permanent Access to the Records of Science. The output from the project is intended to guide the European Commission's strategy about research infrastructure.

Upcoming events

[nestor closing event: 10 June 2009](#)

Recent publications

[Draft Roadmap \(March 2009, PARSE.insight\)](#)

[First insight HEP \(Jan 2009, PARSE.insight\)](#)

[Permanent Access to Raw Scientific Data: PARSE.Insight, D-Lib Magazine](#)